

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 13, 2002, 20:59:48 ; Search time 2883 Seconds

(without alignments)
1655.519 Million cell updates/sec

Title: US-09-659-737A-1

Perfect score: 164

Sequence: 1 ggcacgggaatcaagagcag.....tgccctgagtgcacccagaaag 164

Scoring table: IDENTITY_MDC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb.ba:*

2: gb.htg:*

3: gb.in:*

4: gb.om:*

5: gb.ov:*

6: gb.pat:*

7: gb.ph:*

8: gb.pl:*

9: gb.pr:*

10: gb.ro:*

11: gb.sts:*

12: gb.sy:*

13: gb.un:*

14: gb.vi:*

15: em.ba:*

16: em.fun:*

17: em.hum:*

18: em.in:*

19: em.mu:*

20: em.om:*

21: em.or:*

22: em.ov:*

23: em.pat:*

24: em.ph:*

25: em.pl:*

26: em.ro:*

27: em.sts:*

28: em.un:*

29: em.vi:*

30: em.htg.hum:*

31: em.htg.inv:*

32: em.htg.other:*

33: em.htg.mus:*

34: em.htg.pln:*

35: em.htg.tod:*

36: em.htg.mam:*

37: em.htg.vrt:*

38: em.sy:*

39: em.htgo.hum:*

40: em.htgo.mus:*

41: em.htgo.other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	164	100.0	164	6 AX139125	AX139125 Sequence
2	152.8	93.2	2157	6 AX207410	AX207410 Sequence
3	152.8	93.2	3111	6 AX482007	AX482007 Sequence
4	152.8	93.2	3518	6 AX482009	AX482009 Sequence
5	149.6	91.2	3910	6 HSA311797	AX311797 Homo sapi
6	149.6	91.2	4667	9 HSA311798	AJ311798 Homo sapi
7	136.4	83.2	118632	9 AL133380	AL133380 Human DNA
8	120.8	73.7	3616	10 BC021891	BC021891 Mus muscu
9	106	64.6	168795	2 AC102420	AC102420 Mus muscu
10	104.4	63.7	96101	2 AC112003	AC112003 Rattus no
11	94.2	57.4	3931	9 AF251442	AF251442 Homo sapi
12	83	50.6	3531	9 HSU07747	U07747 Human SH3 d
13	83	50.6	3558	6 AX399680	AX399680 Sequence
14	83	50.6	3558	9 HMMMLK3A	L32976 Human Prote
15	83	50.6	3603	9 BC011263	BC011263 Homo sapi
16	83	50.6	3799	9 AK092015	AK092015 Homo sapi
17	82	50.0	166200	9 AC004816	AC004816 Homo sapi
18	79.2	48.3	39251	5 AF154413	AF154413 Fugu rubr
19	78.8	48.0	197943	2 AC125351	AC125351 Mus muscu
20	78.8	48.0	203574	2 AC124595	AC124595 Mus muscu
21	75	45.7	2390	10 BC030928	BC030928 Mus muscu
22	73.4	44.8	3138	9 HSMSTMR	Z48615 H.sapiens M
23	73.4	44.8	3454	6 AX337846	AX337846 Sequence
24	72.8	44.4	150694	2 HARNAMLK2	X90846 H.sapiens m
25	72.8	44.4	686	11 G33125	AP001459 Homo sapi
26	72.4	44.1	140356	2 AP000803	G33125 Ep10F7a5 Hu
27	72.4	44.1	162445	2 AC022488	AP000803 Homo sapi
28	72.4	44.1	211382	2 AP001362	AC022488 Homo sapi
29	72.4	43.9	404	6 I44517	AP001362 Homo sapi
30	67.8	41.3	66251	2 AC100718	I44517 Sequence 21
31	66.5	40.2	34340	2 AC130251	AC100718 Mus muscu
32	64.8	39.3	168423	2 AC111962	AC130251 Rattus no
33	64.8	39.3	46440	10 AF151142	AC111962 Rattus no
34	62.4	38.0	2139	6 AF151142	AF151142 Mus muscu
35	62.4	37.8	137	6 I44513	AF151142 Mus muscu
36	60	36.5	215615	2 AC074312	AK090614 Homo sapi
37	60	36.5	226059	2 AC079488	I44513 Sequence 13
38	60	36.5	243275	2 AC073705	AC074312 Mus muscu
39	59.8	36.5	140529	2 AC011486	AC079488 Mus muscu
40	59.8	36.5	178335	2 AC118344	AC073705 Mus muscu
41	58.4	35.6	170094	2 AC120811	AC011486 Homo sapi
42	57.4	35.0	3447	3 AF416233	AC118344 Homo sapi
43	57.4	35.0	5129	3 AY045717	AC120811 Rattus no
44	57.4	35.0	5131	3 AY119549	AF416233 Drosophila
45	57.4	35.0	5131	3 AY119549	AY045717 Drosophila

ALIGNMENTS

RESULT 1

AX139125

LOCUS AX139125 164 bp DNA linear PAT 30-MAY-2001

DEFINITION Sequence 1 from Patent EP1085093.

ACCESSION AX139125

VERSION AX139125.1 GI:14274801

KEYWORDS

SOURCE

ORGANISM human.

REFERENCE

AUTHORS Blumenberg, M. and Gazel, A.M.

TITLE Genes and polynucleotides associated with ultraviolet radiation-mediated skin damage and uses thereof

JOURNAL Patent: EP 1085093-A 1 21-MAR-2001;

NEW YORK UNIVERSITY (US)

FEATURES Location/Qualifiers

source

1..164

/organism="Homo sapiens"

/db_xref="taxon:9606"

<2..>163

/note="unnamed protein product"

/codon_start=1

/protein_id="CAC3965.1"

/db_xref="GI:14274802"

/translation="HRDIKAGNILLLEKIEHIDICNKTITDGLAREHRTTKMST

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

AGTYAMWAP"

CDS

BASE COUNT 55 a 33 c 44 g 32 t

ORIGIN

Query Match 100.0%; Score 164; DB 6; Length 164;

Best Local Similarity 100.0%; Pred. No. 4, 4e-38;

Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCACCGGACATCAAGCGAGGAATATTTGCTACTTGAGAGATAGAACATGATGACAT 60
|||||
Db 1 GCACCGGACATCAAGCGAGGAATATTTGCTACTTGAGAGATAGAACATGATGACAT 60QY 61 CTGCAATPAAACTTTGAAGATTACAGATTTGGGTTGGCGAGGAATGGCACAGACCAC 120
|||||
Db 61 CTGCAATPAAACTTTGAAGATTACAGATTTGGGTTGGCGAGGAATGGCACAGACCAC 120QY 121 CAAATATGAGCAGACGACACCTATGCTGATGGCCCCAGAG 164
|||||
Db 121 CAAATATGAGCAGACGACACCTATGCTGATGGCCCCAGAG 164RESULT 2
AX207410 2157 bp DNA linear PAT 30-AUG-2001

LOCUS Sequence 23 from Patent WO0155356.

DEFINITION AX207410

ACCESSION AX207410.1 GI:15395228

VERSION

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2157)

AUTHORS Plozman,G., Whyte,D., Manning,G., Sudarsanam,S. and Martinez,R.

TITLE Human protein kinases and protein kinase-like enzymes

JOURNAL Patent: WO 0155356-A 23 02-AUG-2001;

SUGEN, Inc. (US)

FEATURES Location/Qualifiers

1..2157

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 514 a 611 c 638 g 394 t

ORIGIN

Query Match 93.2%; Score 152.8; DB 6; Length 2157;
Best Local Similarity 95.7%; Pred. No. 1, 1e-34;
Matches 157; Conservative 0; Mismatches 7; Indels 0; Gaps 0;QY 1 GCACCGGACATCAAGCGAGGAATATTTGCTACTTGAGAGATAGAACATGATGACAT 60
|||||
Db 780 GCACCGGACATCAAGCGAGGAATATTTGCTACTTGAGAGATAGAACATGATGACAT 839
QY 61 CTGCAATPAAACTTTGAAGATTACAGATTTGGGTTGGCGAGGAATGGCACAGACCAC 120
|||||
Db 840 CTGCAATPAAACTTTGAAGATTACAGATTTGGGTTGGCGAGGAATGGCACAGACCAC 899
QY 121 CAAATATGAGCAGACGACACCTATGCTGATGGCCCCAGAG 164
|||||
Db 900 CAAATATGAGCAGACGACACCTATGCTGATGGCCCCAGAG 943RESULT 3
AX482007 3111 bp DNA linear PAT 16-AUG-2002

LOCUS Sequence 1 from Patent WO02055685.

DEFINITION AX482007

ACCESSION AX482007.1 GI:22316731

VERSION

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1

AUTHORS Hu,Y., Kieke,J.A. and Donoho,G.

TITLE Novel human kinase and polynucleotides encoding the same

JOURNAL Patent: WO 02055685-A 1 18-JUL-2002;

Lexicon Genetics Incorporated (US)

FEATURES Location/Qualifiers

1..3111

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 740 a 905 c 864 g 594 t 8 others

ORIGIN

Query Match 93.2%; Score 152.8; DB 6; Length 3111;
Best Local Similarity 95.7%; Pred. No. 1, 1e-34;
Matches 157; Conservative 0; Mismatches 7; Indels 0; Gaps 0;QY 1 GCACCGGACATCAAGCGAGGAATATTTGCTACTTGAGAGATAGAACATGATGACAT 60
|||||
Db 780 GCACCGGACATCAAGCGAGGAATATTTGCTACTTGAGAGATAGAACATGATGACAT 839QY 61 CTGCAATPAAACTTTGAAGATTACAGATTTGGGTTGGCGAGGAATGGCACAGACCAC 120
|||||
Db 61 CTGCAATPAAACTTTGAAGATTACAGATTTGGGTTGGCGAGGAATGGCACAGACCAC 899QY 121 CAAATATGAGCAGACGACACCTATGCTGATGGCCCCAGAG 164
|||||
Db 900 CAAATATGAGCAGACGACACCTATGCTGATGGCCCCAGAG 943RESULT 4
AX482009 3518 bp DNA linear PAT 16-AUG-2002

LOCUS Sequence 3 from Patent WO02055685.

DEFINITION AX482009

ACCESSION AX482009.1 GI:22316732

VERSION

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1

AUTHORS Hu,Y., Kieke,J.A. and Donoho,G.

TITLE Novel human kinase and polynucleotides encoding the same

JOURNAL Patent: WO 02055685-A 3 18-JUL-2002;

Lexicon Genetics Incorporated (US)

FEATURES Location/Qualifiers

1..3518

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 811 a 1043 c 976 g 680 t 8 others

ORIGIN

Query Match 93.2%; Score 152.8; DB 6; Length 3518;
Best Local Similarity 95.7%; Pred. No. 1, 1e-34;
Matches 157; Conservative 0; Mismatches 7; Indels 0; Gaps 0;QY 1 GCACCGGACATCAAGCGAGGAATATTTGCTACTTGAGAGATAGAACATGATGACAT 60
|||||
Db 1043 GCACCGGACATCAAGCGAGGAATATTTGCTACTTGAGAGATAGAACATGATGACAT 1102
QY 61 CTGCAATPAAACTTTGAAGATTACAGATTTGGGTTGGCGAGGAATGGCACAGACCAC 120
|||||

Db 1103 CTGCATTAACCTTTGAAGATTACAGATTTTGGGTTGGCGAGGGAATGCCACAGACAC 1162

QY 121 CAAATATGACACAGCAGCAGCCTATGCTGATGGCCCCGAGAG 164

Db 1163 CAAATATGACACAGCAGCAGCCTATGCTGATGGCCCCGAGAG 1206

RESULT 5

HS311797

LOCUS HSA311797 3910 bp mRNA linear PRI 12-DEC-2001

DEFINITION Homo sapiens mRNA for mixed lineage kinase 4alpha (MLK4ALPHA gene).

ACCESSION AJ311797

VERSION AJ311797.1 GI:17736728

KEYWORDS mixed lineage kinase 4alpha; MLK4ALPHA gene.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1

AUTHORS Kashuba, V., Protodopov, A., Rynditch, A., Zabarovsky, E. and Kashuba, V.

TITLE MLK4, a new member of mixed lineage kinases

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 3910)

AUTHORS Kashuba, V.

TITLE Direct Submission

JOURNAL Submitted (12-APR-2001) Kashuba V., Microbiology and Tumorbiology Center, Karolinska Institute, Box 280, 171 77, SWEDEN

FEATURES

source

1. .3910

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="1"

/map="1q42"

/tissue_type="heart"

262. .1974

/gene="MLK4ALPHA"

262. .1974

/db_xref="GI:17736728"

/translation="MALRGAAGATPTPVSSAGAPGSGASSSTSGSASAGAGLMA

ALDYEARGEDELSTRGOLIVEYLSDAAVSGDEGMAGOVORLGIPTANYAPCRP

AASPAPEPSPSPVHYAFERLELKEKELIGAGGCGYRATNOGEVAYAKAROPBD

AAAABESVREARLEFAMLRHPNITELRGVCLQDPHLCVLEFARGALNRLAANA

PDPARPGRRARRIPPHVLYNNAVOIARGLYDHEAFVPIILRDLKSSNILLLEIE

HDIDCNKTLKIDFGELAREMHRTTMSAGTAYAMAPVIVKSLFSGSDIWSGVIL

WELLTGEVYRIGIDGLAVAGVAVNKLTLPIPTCPEPEFAKMECMOODPHIRPSFA

LLEIOLTAIEGAVMTEMPOESHMSODDKLEIOMDELTKEKELRSREELTRAA

LOOKSOBELKREDOALAREIDYLERELNILLIOLNOEKRYKRGKFRSRILK

DGHRISLPDSFOHKITVOASPNIDKRSLNSSSSPSPSTMPRLRAIOCELALPLP

GLIC"

BASE COUNT 923 a 1038 c 1056 g 893 t

polyA-signal 3893. .3898

ORIGIN

Query Match 91.2%; Score 149.6; DB 9; Length 3910;

Best Local Similarity 94.5%; Pred. No. 1e-33;

Matches 155; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GCACCGGAGACATCAAGGAGGAATATTTGCTACTTGAGAGATGAAACATGATGACAT 60

Db 1041 GCACCGGAGACCTCAAGTCCACACATTTTGCTACTTGAGAGATGAAACATGATGACAT 1100

QY 61 CTGCATTAACCTTTGAAGATTACAGATTTTGGGTTGGCGAGGGAATGCCACAGACAC 120

Db 1101 CTGCATTAACCTTTGAAGATTACAGATTTTGGGTTGGCGAGGGAATGCCACAGACAC 1160

QY 121 CAAATATGACACAGCAGCAGCCTATGCTGATGGCCCCGAGAG 164

Db 1161 CAAATATGACACAGCAGCAGCCTATGCTGATGGCCCCGAGAG 1204

RESULT 6

HS311798

LOCUS HSA311798 4667 bp mRNA linear PRI 12-DEC-2001

DEFINITION Homo sapiens mRNA for mixed lineage kinase 4beta (MLK4BETA gene).

ACCESSION AJ311798

VERSION AJ311798.1 GI:17736730

KEYWORDS mixed lineage kinase 4beta; MLK4BETA gene.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1

AUTHORS Kashuba, V., Protodopov, A., Rynditch, A., Zabarovsky, E. and Kashuba, V.

TITLE MLK4, a new member of mixed lineage kinases

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 4667)

AUTHORS Kashuba, V.

TITLE Direct Submission

JOURNAL Submitted (12-APR-2001) Kashuba V., Microbiology and Tumorbiology Center, Karolinska Institute, Box 280, 171 77, SWEDEN

FEATURES

source

1. .4667

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="1"

/map="1q42"

262. .3372

/gene="MLK4BETA"

262. .3372

/db_xref="GI:17736731"

/translation="MALRGAAGATPTPVSSAGAPGSGASSSTSGSASAGAGLMA

ALDYEARGEDELSTRGOLIVEYLSDAAVSGDEGMAGOVORLGIPTANYAPCRP

AASPAPEPSPSPVHYAFERLELKEKELIGAGGCGYRATNOGEVAYAKAROPBD

AAAABESVREARLEFAMLRHPNITELRGVCLQDPHLCVLEFARGALNRLAANA

PDPARPGRRARRIPPHVLYNNAVOIARGLYDHEAFVPIILRDLKSSNILLLEIE

HDIDCNKTLKIDFGELAREMHRTTMSAGTAYAMAPVIVKSLFSGSDIWSGVIL

WELLTGEVYRIGIDGLAVAGVAVNKLTLPIPTCPEPEFAKMECMOODPHIRPSFA

LLEIOLTAIEGAVMTEMPOESHMSODDKLEIOMDELTKEKELRSREELTRAA

LOOKSOBELKREDOALAREIDYLERELNILLIOLNOEKRYKRGKFRSRILK

DGHRISLPDSFOHKITVOASPNIDKRSLNSSSSPSPSTMPRLRAIOELSDENK

TGKRTVFRQEEFEEDVKNRNFKCGCTWPNISJOMKRDCKERRAPLSDGNSPSTIL

INOKTMEPLASLFEVQPSCEEPKLSPDGLSHRKRQIKLPSQAVIDLPLGDKDQREN

PEAGSMEEAASANAATYTIEMAPNLSRSRPOKRTESALYGCIVLAVALGIDLR

ELHKRQAAEEPLPREKREKREIIPQRAKSRSPASPSISSTGSEASSPSSLASA

LCILTPSTSTCILLQMSSEDLVDSAYTCDSEMLTIDFPCPTAPGSRERPALPRUD

TDCSVSRNLPSSEFLQRTGNNVYCASSHRBSHHRRTSDSNLPTPTGATLISATGASA

LPICSPAPPHSHLPREVSIPKSHSVHAIIVOPRPAISRSDLPQAVPQTAVSOALQTA

CVGGRPGPHPTQFLAAKREKTSHPVSLLDVVEGOSRDYTVPLGRMRKSTRPSIYEL

EKEFLS"

BASE COUNT 1153 a 1228 c 1167 g 1119 t

ORIGIN

Query Match 91.2%; Score 149.6; DB 9; Length 4667;

Best Local Similarity 94.5%; Pred. No. 1e-33;

Matches 155; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GCACCGGAGACATCAAGGAGGAATATTTGCTACTTGAGAGATGAAACATGATGACAT 60

Db 1041 GCACCGGAGACCTCAAGTCCACACATTTTGCTACTTGAGAGATGAAACATGATGACAT 1100

QY 61 CTGCATTAACCTTTGAAGATTACAGATTTTGGGTTGGCGAGGGAATGCCACAGACAC 120

Db 1101 CTGCATTAACCTTTGAAGATTACAGATTTTGGGTTGGCGAGGGAATGCCACAGACAC 1160

QY 121 CAAATATGACACAGCAGCAGCCTATGCTGATGGCCCCGAGAG 164

|||||

Db 1161 CAATATGACACAGCAGCCTATGCCTGATGGCCCCGAG 1204

RESULT 7
AL133380
LOCUS
DEFINITION

AL133380 118632 bp DNA linear PRI 29-NOV-2000
Human DNA sequence from clone RP5-862P8 on chromosome 1q42.2-43
Contains a 40S ribosomal protein S7 pseudogene, start of a gene for
a novel protein (similar to C.elegans and D.melanogaster), a gene
similar to MAP3K10 (mitogen-activated protein kinase kinase
10), STS, GSSS and a Cpg island, complete sequence.

AL133380
VERSION
KEYWORDS
SOURCE
ORGANISM

AL133380.5 GI:8217443
HTG: Cpg island; protein kinase; Ribosomal protein.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 118632)

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (31-MAY-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 3, 2000 this sequence version replaced gi:7798842.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chri
RP5-862P8 is from the library RPCI-5 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2

FEATURES
source

This sequence is the entire insert of clone RP5-862P8.
Location/Qualifiers
1..118632
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q42.2-43"
/clone="RP5-862P8"
/clone_lib="RPCI-5"
323..682
repeat_region
/note="THE1B repeat: matches 1..362 of consensus"
742..953
repeat_region
/note="MIR repeat: matches 45..262 of consensus"
misc_feature
complement(1784..2072)
/note="match: GSS: Em:AQ887379"
misc_feature
complement(1793..1994)
/note="match: GSS: Em:AQ384546"
1793..1850
repeat_region
/note="29 copies 2 mer tt 74% conserved"
misc_feature
complement(1810..2413)
/note="match: SNS: Em:D11774"
misc_feature
complement(1812..1928)

gene
CDS

/note="match: GSS: Em:AQ219188"
complement(1849..2433)
/gene="dJ862P8.1"
complement(<1849..2433)
/gene="dJ862P8.1"
/note="dJ862P8.1 (40S Ribosomal Protein S7 pseudogene)
match: CDNAS: Em:X74803 Em:U16258 Em:V01442 Em:M77233
Em:U18487
match: ESTs: Em:AI707735
match: proteins: Sw:P02362 Sw:P23821 Sw:P50894"
/pseudo
/codon_start=1
/evidence=not_experimental
1873..2009
misc_feature
/note="match: GSS: Em:AQ476700"
1875..2308
misc_feature
/note="match: GSS: Em:AQ593042"
2153..2376
misc_feature
/note="match: GSS: Em:AQ264373"
2157..2439
misc_feature
/note="match: GSS: Em:AQ554409"
2186..2451
/note="match: SNS: Em:U16258"
complement(2283..2360)
/gene="dJ862P8.1"
/note="match: SNS: Em:H84775"
2841..2910
repeat_region
/note="MER91A repeat: matches 1..69 of consensus"
4083..4220
repeat_region
/note="MIR repeat: matches 2..142 of consensus"
4207..4313
repeat_region
/note="MIR repeat: matches 62..173 of consensus"
4314..4623
repeat_region
/note="Alusp repeat: matches 1..312 of consensus"
5181..5494
repeat_region
/note="Alusx repeat: matches 1..312 of consensus"
complement(6410..6697)
/note="match: GSS: Em:AQ088034"
/note="match: GSS: Em:U16258"
/note="match: GSS: Em:AQ177047"
6684..6864
repeat_region
/note="MIR repeat: matches 1..247 of consensus"
6974..7178
repeat_region
/note="MIR repeat: matches 34..256 of consensus"
7203..7407
repeat_region
/note="MIR repeat: matches 2..208 of consensus"
complement(8312..8494)
/gene="dJ862P8.2"
complement(<8312..8494)
/gene="dJ862P8.2"
/product="dJ862P8.2 (5' end of gene encoding novel protein
similar to C.elegans and D.melanogaster)"
/evidence=not_experimental
complement(<8312..8464)
/gene="dJ862P8.2"
/note="match: proteins: Tr:O46074"
/note="match: proteins: Tr:O46074"
/codon_start=1
/evidence=not_experimental
/product="dJ862P8.2 (5' end of gene encoding novel protein
similar to C.elegans and D.melanogaster)"
/protein_id="CAC17570.1"
/db_xref="GI:11493212"
/translation="MVSQVQLRLRQGVMAALNGWYHDPQSKFTNSCHLYLMLFLL
LPRLHL"
8840..9031
repeat_region
/note="96 copies 2 mer ct 70% conserved"
10240..10362
repeat_region
/note="L2 repeat: matches 2576..2702 of consensus"
10558..10791
repeat_region
/note="L2 repeat: matches 2157..2417 of consensus"
11593..11897
repeat_region
/note="Alusx repeat: matches 1..312 of consensus"
11942..12373

```

repeat_region      /note="L2 repeat: matches 1642. .2091 of consensus"
12715. .12791
/note="LIR33 repeat: matches 99. .175 of consensus"
repeat_region      13110. .13274
/note="LIM4 repeat: matches 4431. .4608 of consensus"
misc_feature        14285. .14806
/note="match: GSS: Em:AQ476745"
repeat_region      14369. .14623
/note="LIM4 repeat: matches 3588. .3847 of consensus"
repeat_region      14684. .14781
/note="HAL1 repeat: matches 842. .935 of consensus"
repeat_region      14782. .15078
/note="AlusX repeat: matches 1. .300 of consensus"
repeat_region      15079. .15181
/note="HAL1 repeat: matches 935. .1051 of consensus"
repeat_region      15294. .15791
/note="LIM4 repeat: matches 2865. .3375 of consensus"
repeat_region      16096. .16185
/note="LIM4 repeat: matches 2693. .2784 of consensus"
repeat_region      16241. .16506
/note="Alub repeat: matches 24. .288 of consensus"
repeat_region      16509. .16548
/note="20 copies 2 mer at 77% conserved"
repeat_region      16577. .16638
/note="LIM4 repeat: matches 2375. .2437 of consensus"
repeat_region      16639. .16726
/note="LIP4 repeat: matches 6057. .6144 of consensus"
repeat_region      16727. .16989
/note="LIM4 repeat: matches 2120. .2375 of consensus"
repeat_region      16742. .17063
/note="LIMC repeat: matches 2407. .2364 of consensus"
repeat_region      18214. .18422
/note="LIMC repeat: matches 2105. .2311 of consensus"
repeat_region      18423. .18740
/note="Alub repeat: matches 1. .309 of consensus"
repeat_region      18741. .19382
/note="LIMC repeat: matches 2311. .2620 of consensus"
repeat_region      19580. .19641
/note="LIMB repeat: matches 6113. .6174 of consensus"
repeat_region      19776. .19948
/note="LIME3A repeat: matches 5980. .6160 of consensus"
repeat_region      20173. .20579
/note="MSTB repeat: matches 1. .422 of consensus"
repeat_region      20631. .20842
/note="MIR repeat: matches 22. .245 of consensus"
repeat_region      21161. .21474
/note="AlusX repeat: matches 1. .312 of consensus"
repeat_region      21772. .21873
/note="51 copies 2 mer aa 60% conserved"
repeat_region      22246. .22531
/note="AlusX repeat: matches 1. .289 of consensus"
repeat_region      22582. .22894
/note="AlusX repeat: matches 1. .311 of consensus"
repeat_region      23101. .23421
/note="LIM4 repeat: matches 2740. .3046 of consensus"
repeat_region      23468. .23548
/note="MIR1J repeat: matches 104. .191 of consensus"
repeat_region      23643. .24023
/note="MSTB repeat: matches 1. .418 of consensus"
repeat_region      24596. .24676

```

```

Query Match      83.2% Score 136.4; DB 9; Length 118632;
Best Local Similarity 99.3%; Pred. No. 1.1e-29;
Matches 137; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 27 TTTTGCTACTTGAGAGATAGAACATGATGACATCTGCAATTAACCTTGAAGATTACAG 86
|||||
Db 59428 TTTTGCTACTTGAGAGATAGAACATGATGACATCTGCAATTAACCTTGAAGATTACAG 59487
|||||
Oy 87 ATTTTGGCTTGGCGAGGAATGGACAGACACCAAAATGAGCAGACAGGACCTTATG 146
|||||
Db 59488 ATTTTGGCTTGGCGAGGAATGGACAGACACCAAAATGAGCAGACAGGACCTTATG 59547
|||||

```

```

Oy 147 CCTGATGGCCCCAGAG 164
Db 59548 CCTGATGGCCCCAGAG 59565

```

```

RESULT 8
BC021891
LOCUS
DEFINITION
Mus musculus, similar to mitogen-activated protein kinase
kinase 9, clone MGC:27778 IMAGE:3156324, mRNA, complete cds.
ACCESSION
BC021891
VERSION
BC021891.1 GI:18257337
KEYWORDS
MGC.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE
1 (bases 1 to 3616)
AUTHORS
Strausberg,R.
TITLE
Submitted (18-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

```

```

REMARK
COMMENT
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Huilyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>
Series: IRAC Plate: 35 Row: d Column: 15
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

```

FEATURES
source
location/Qualifiers
1..3616
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="MGC:27778 IMAGE:3156324"
/tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."
/clone_lib="NCL_CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
224..3229
/codon_start=1
/product="Similar to mitogen-activated protein kinase
kinase 9"
/protein_id="AAH21891.1"
/db_xref="GI:18257338"
/translat="MALPYAEGTADPLSPARDGSGTSSGMAALYDEARGEDELS
LRGQLEVEVLSODAAVSGDGMAGOVORRLGIFPASVYAPGPPAPPPRCSIP
VHDFPERLEKELIGAGGGGOYRATVMOGOEYAVKARRDPODAAAEVRRARL
FALRPNTIQLRGVCLRQPHCTIVFARGALNKAALAAASDPRAPRRARRIP
QAVLVNAVOJARGMILHBEAAYVPIIHRDKSNITLLEKIDIDICNKTITPGL
ARWHRTPRASAGTYAWMAPEVTRSLTSSKSDTWSTGLVIMELTGEVPRGIDGL
AVAYGAVNKLTLPIPTCPEPPAKLMEKCSWDDPIRPSFALIILQULVAIEAVLTN
MPESEFHSMDKRLKLEIOMFSELTAKELKSEDEELISPAALQOQSOELLRLRQDL
AEREIVLERELNVLIFOLSEAPHYVKKRGRFRRLKDGHRILSPSDQKHTV
QASPTIDKRSSSDSGICSPGSPIMLPRILATOLITDENNKTRGRNMYVRODPEDVK
RSFKKKGCTMGPSVOTKERPERREVRPLSDGNSWSSLILKSOKTTPLAFLVDOP

```

CDS

```

GCSEKUEVLEGELEHRRKOTKPPGQAHVGLPLCKDSOREDDSPARESPKSPY
NNGAPMLRKKTESALCEGCMILASMALGIDRYLHGAQAPAKSPMEKKEBALDP
ASRQSSPSLLINPSPKGRAPSGSSTILLPSAPSHSKSSLSIMCLLOAGEESLIDP
ARLDCGTTTTPDPGASAPSGELLPLGLPKDYGLVLRPMHAILDGTGERLPGCAI
VGDKCHHMQMGSSEETPLWLOSAPEBDSPLHSPSPGQRIASQASLVKEGVGECQ
ACALPDQRPTASVRTSPPTWCDKHQVPAALACLLGAQERSRCQPPSLDASIEGO
KKQCAMPLCHVKSMCPSIYALEKDELTA"

BASE COUNT      813 a 1036 c 1074 g 693 t
ORIGIN

Query Match      73.7%; Score 120.8; DB 10; Length 3616;
Best Local Similarity 83.5%; Pred. No. 3.5e-25;
Matches 137; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 1 GCACGGGACATCAAGCAGCAATATTTGCTACTTGAGAGATAGACATGATGACAT 60
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 955 GCACCGGACGCTCAAGTCACGACACATTTCTGCTGAGAGATAGACATGATGACAT 1014
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 61 CTGCAATPAAACTTGTGAAGATTACAGATTTTGGGTCGCGAGGAGATGCGACAGACCAC 120
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1015 TTGCAACAGACGCTGGAAGATCAGGACTTTGGGCTAGCGAGGAGATGCGACAGACCAC 1074
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 121 CAAATATGAGCAGCAGCAGCAGCCTATGCTGATGATGCCCCCAGAG 164
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1075 CAGGATAGCGCGACGACGAGCCTAGCCTGATGCTGCCGAGG 1118
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
AC102420      168795 bp      DNA      linear      HTG 21-AUG-2002
LOCUS
DEFINITION Mus musculus clone RP24-553021, WORKING DRAFT SEQUENCE, 25
unordered pieces.
AC102420      GI:22380990
VERSION
AC102420.2
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 168795)
REFERENCE Birren, B., Nusbaum, C. and Lander, E.
AUTHORS Mus musculus, clone RP24-553021
JOURNAL Unpublished
TITLE
JOURNAL
REFERENCE
AUTHORS

2 (bases 1 to 168795)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barrera, N., Bastien, V., Boguslavsky, L., Bouhagalter, B.,
Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Labroque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrim, J.,
Meneus, J., Minova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodores, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 168795)
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barrera, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouhagalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,

```

TITLE
JOURNAL
COMMENT

```

Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Fero, S., Ferrelia, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, J., Minova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Testaye, S., Theodores, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:17061506.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L18760
Center clone name: 553_O-21
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161900 bases at least Q40
Consensus quality: 165140 bases at least Q30
Consensus quality: 166006 bases at least Q20
Insert size: 176000; agarose-ef
Insert size: 166395; sum-of-ctrls
Quality coverage: 5.7 in Q20 bases; agarose-ef
Quality coverage: 6.1 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
895 994: gap of 100 bp in length
995 2169: contig of 1175 bp in length
2170 2269: gap of 100 bp
2270 3068: contig of 799 bp in length
3069 3168: gap of 100 bp
3169 4507: contig of 1339 bp in length
4508 4607: gap of 100 bp
4608 5627: contig of 1020 bp in length
5628 5727: gap of 100 bp
5728 7115: contig of 1388 bp in length
7116 7215: gap of 100 bp
7216 8963: contig of 1748 bp in length
8964 9063: gap of 100 bp
9064 11172: contig of 2649 bp in length
11713 11812: gap of 100 bp
11813 14077: contig of 2265 bp in length
14078 14177: gap of 100 bp
14178 17221: contig of 3044 bp in length
17222 17321: gap of 100 bp
17322 19510: contig of 2189 bp in length
19511 19610: gap of 100 bp
19611 23807: contig of 4197 bp in length

```

*	23808	23907:	gap of	100	bp
*	23908	26130:	contig of 2223	bp	in length
*	26131	26230:	gap of	100	bp
*	26231	28974:	contig of 2744	bp	in length
*	28975	29074:	gap of	100	bp
*	29075	31408:	contig of 2334	bp	in length
*	31409	31508:	gap of	100	bp
*	31509	33532:	contig of 3844	bp	in length
*	33533	35452:	gap of	100	bp
*	35453	40317:	contig of 4865	bp	in length
*	40318	40417:	gap of	100	bp
*	40418	47500:	contig of 7083	bp	in length
*	47501	47600:	gap of	100	bp
*	47601	54652:	contig of 7052	bp	in length
*	54653	54752:	gap of	100	bp
*	54753	67192:	contig of 12440	bp	in length
*	67193	67292:	gap of	100	bp
*	67293	80111:	contig of 12819	bp	in length
*	80112	80211:	gap of	100	bp
*	80212	94967:	contig of 14756	bp	in length
*	94968	95067:	gap of	100	bp
*	95068	116109:	contig of 21042	bp	in length
*	116110	116209:	gap of	100	bp
*	116210	137004:	contig of 20795	bp	in length
*	137005	137104:	gap of	100	bp
*	137105	168795:	contig of 31691	bp	in length

URES	Location/Qualifiers
source	1. .168795
	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/clone="RP24-353021"
	/clone_lib="RPC1-24 Male Mouse BAC"
misc_feature	1. .894
	/note="assembly_fragment"
misc_feature	995. .2169
	/note="assembly_fragment"
misc_feature	2270. .3068
	/note="assembly_fragment"
misc_feature	3169. .4507
	/note="assembly_fragment"
misc_feature	4608. .5627
	/note="assembly_fragment"
misc_feature	5728. .7115
	/note="assembly_fragment"
misc_feature	7216. .8963
	/note="assembly_fragment"
misc_feature	9064. .11712
	/note="assembly_fragment"
misc_feature	11813. .14077
	/note="assembly_fragment"
misc_feature	14178. .17221
	/note="assembly_fragment"
misc_feature	17322. .19510
	/note="assembly_fragment"
misc_feature	19611. .23807
	/note="assembly_fragment"
misc_feature	23908. .26130
	/note="assembly_fragment"
misc_feature	26231. .28974
	/note="assembly_fragment"
misc_feature	29075. .31408
	/note="assembly_fragment"
misc_feature	31509. .35352
	/note="assembly_fragment"
misc_feature	35453. .40317
	/note="assembly_fragment"
misc_feature	40418. .47500
	/note="assembly_fragment"
misc_feature	47601. .54652
	/note="assembly_fragment"
misc_feature	54753. .67192
	/note="assembly_fragment"
misc_feature	67293. .80111

[illegible]

LOCUS	AC112003/C	RESULT 10
DEFINITION	AC112003	96101 bp DNA
ACCESSION	AC112003	linear
VERSION	AC112003.2	HTG 13-JUL-2002
KEYWORDS	HTG; HTGS; PHASE1.	SEQUENCING IN PROGRESS
SOURCE	Norway rat.	
ORGANISM	Rattus norvegicus	
	Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
REFERENCE	1 (bases 1 to 96101)	
AUTHORS	Munzy,D.M., Adams,C., Adlo-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,K.M., Banks,T., Barbata,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowle,S., Britva,M., Brown,E., Brown,M., Bryant,N.P., Bunah,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N., C., Carron,T.F., Carter,M., Cavazos,C.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denna,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durlin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,S., Escoto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frintz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolyvet,S., Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratochvic,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissegod,H., Lozdo,R.U., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., Mcleod,M.P., Meador,M., Mel,G., Metzler,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,S., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokewko,S., Oguh,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojudoxani,I., Rolfe,M., Ruiz,S., Saverly,G.,	

Scherer, S., Scott, G., Shen, H., Shooshtrani, N., Sisson, I.,
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S.,
Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 96101)
Worley, K.C.

Direct Submission
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 96101)
Worley, K.C.

Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18701953.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc.help@bcm.tmc.edu

----- Project Information
Center project name: G02X
Center clone name: CH230-234D24

----- Summary Statistics
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 49084 bases at least Q40
Consensus quality: 53702 bases at least Q30
Consensus quality: 57526 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 49 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1114: contig of 1114 bp in length
* 1115 1214: gap of unknown length
* 1215 2722: contig of 1508 bp in length
* 2723 2822: gap of unknown length
* 2823 4182: contig of 1360 bp in length
* 4183 4282: gap of unknown length
* 4283 5938: contig of 1656 bp in length
* 5939 6038: gap of unknown length
* 6039 7047: contig of 1009 bp in length
* 7048 7147: gap of unknown length
* 7148 8409: contig of 1262 bp in length
* 8410 8509: gap of unknown length
* 8510 9889: contig of 1380 bp in length
* 9890 9990: gap of unknown length
* 9990 11462: contig of 1473 bp in length
* 11463 11563: gap of unknown length
* 11563 12742: contig of 1180 bp in length
* 12743 12842: gap of unknown length
* 12843 13851: contig of 1009 bp in length
* 13852 13952: gap of unknown length
* 13952 15437: contig of 1486 bp in length
* 15438 15537: gap of unknown length
* 15538 17126: contig of 1589 bp in length

17127 17226: gap of unknown length
* 17227 18632: contig of 1406 bp in length
* 18633 18732: gap of unknown length
* 18733 20346: contig of 1614 bp in length
* 20347 20447: gap of unknown length
* 20447 21605: contig of 1159 bp in length
* 21606 21705: gap of unknown length
* 21706 22949: contig of 1244 bp in length
* 22950 24436: gap of unknown length
* 24437 24536: gap of unknown length
* 24537 26037: contig of 1501 bp in length
* 26038 26137: gap of unknown length
* 26138 28251: contig of 2124 bp in length
* 28252 29946: contig of 1585 bp in length
* 28362 30046: gap of unknown length
* 29947 31352: contig of 1306 bp in length
* 30047 31353: gap of unknown length
* 31353 32528: contig of 1076 bp in length
* 31453 32528: gap of unknown length
* 32529 34558: contig of 1930 bp in length
* 32629 34658: gap of unknown length
* 34659 36487: contig of 1829 bp in length
* 36488 36587: gap of unknown length
* 36588 38910: contig of 2323 bp in length
* 38911 39010: gap of unknown length
* 39011 40486: contig of 1476 bp in length
* 40487 40586: gap of unknown length
* 40587 41855: contig of 1268 bp in length
* 41855 41955: gap of unknown length
* 41955 44349: contig of 2395 bp in length
* 44350 44449: gap of unknown length
* 44450 46817: contig of 2368 bp in length
* 46818 46917: gap of unknown length
* 46917 48441: contig of 1524 bp in length
* 48441 48541: gap of unknown length
* 48542 49639: contig of 1098 bp in length
* 49640 49740: gap of unknown length
* 49740 51158: contig of 1419 bp in length
* 51159 51258: gap of unknown length
* 51159 53003: contig of 1745 bp in length
* 51259 53103: gap of unknown length
* 53104 55784: contig of 2661 bp in length
* 55785 55884: gap of unknown length
* 55885 58448: contig of 2564 bp in length
* 58449 58548: gap of unknown length
* 58549 60250: contig of 1702 bp in length
* 60251 60350: gap of unknown length
* 60351 62642: contig of 2292 bp in length
* 62643 62742: gap of unknown length
* 62743 63903: contig of 1161 bp in length
* 63904 64003: gap of unknown length
* 64004 66598: contig of 2595 bp in length
* 66599 66698: gap of unknown length
* 66699 69090: contig of 2392 bp in length
* 69091 69190: gap of unknown length
* 69191 72504: contig of 3414 bp in length
* 72505 72704: gap of unknown length
* 72705 75514: contig of 2810 bp in length
* 75515 75614: gap of unknown length
* 75615 78089: contig of 2475 bp in length
* 78089 78189: gap of unknown length
* 78190 79571: contig of 1382 bp in length
* 79572 79671: gap of unknown length
* 79672 82016: contig of 2345 bp in length
* 82017 82116: gap of unknown length
* 82117 84149: contig of 2033 bp in length
* 84150 84249: gap of unknown length
* 84250 88193: contig of 3944 bp in length
* 88194 88293: gap of unknown length
* 88294 91731: contig of 3438 bp in length
* 91732 91831: gap of unknown length

FEATURES	*	91832	96101:	config of 4270 bp in length.	
SOURCE				Location/Qualifiers	
			1..96101		
			/organism="Rattus norvegicus"		
			/db_xref="taxon:10116"		
			/clone="CH230-234D24"		
BASE COUNT	23908 a	21808 c	22462 g	23062 t	4861 others
ORIGIN					
Query Match		63.7%;	Score 104.4;	DB 2;	Length 96101;
Best Local Similarity	84.8%;	Pred. No. 3.2e-20;			
Matches 111;	Conservative	0;	Mismatches	21;	Indels 0;
			Gaps	0;	
Oy	27	TTTTGCTACTGTGAGAAAGTATGACATGATGATCATCTGCATTAACAACTTTGATGAGATTACAG	86		
Db	33910	TTTTGCTCTCTGGAGAAAGTATGACATGATGATCATCTTGCAACAAACCTTGAGATTCACGG	33911		
Oy	87	ATTTTGGTTGGCGAGGAAATGTCACAGACACCAAAATGATGACACAGCAGACCTATG	146		
Db	33910	ACTTGCGGTTGGCGAGGAAATGTCACAGACACCAAGATGATGACGACAGCAAGACCTAAG	33851		
Oy	147	CCTGGATGGCCCCAGAAC	164		
Db	33850	CCTGGATGGCTCTGAGAG	33833		
RESULT 11					
LOCUS	AF251442				
DEFINITION	Homo sapiens mixed lineage kinase MLK1 mRNA, partial cds.				
ACCESSION	AF251442				
VERSION	AF251442.1				
KEYWORDS	GI:12005723				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens.				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
REFERENCE	1 (bases 1 to 3931)				
AUTHORS	McNee,J.J., Dower,S.K. and Guesdon, F.				
JOURNAL	cDNA sequence and gene organisation of mixed lineage kinase 1				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 3931)				
TITLE	McNee,J.J. and Guesdon, F.				
JOURNAL	Direct Submission				
REFERENCE	Submitted (03-APR-2000) Molecular and Genetic Medicine, University				
AUTHORS	of Sheffield, Royal Hallamshire Hospital, Glossop Road, Sheffield				
JOURNAL	S10 2UF, United Kingdom				
FEATURES					
SOURCE	Location/Qualifiers				
	1..3931				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	<1..3202				
	/note="protein kinase"				
	/codon_start=2				
	/product="mixed lineage kinase MLK1"				
	/protein_id="AA044591.1"				
	/db_xref="GI:12005724"				
	/translation="PLPYWIAVFEYEAAGEDLTLRGLGDVYEVLSKDSQVSGDEGWT				
	GLNRAGIATION-"PLPYWIAVFEYEAAGEDLTLRGLGDVYEVLSKDSQVSGDEGWT				
	GOLNRAGIATION-"PLPYWIAVFEYEAAGEDLTLRGLGDVYEVLSKDSQVSGDEGWT				
	EGKRYAFWIDEEVAAYRAARHDPDEDISQRTENYRQEAFLAMKHENIYALRGVCLK				
	EPNCLIVAEFARGGPIRLNVRISGRKIPDIDLVNMAVVOIARGMVYLAHDAIYDIIHRDLK				
	SNLTILOKVENGLSNKTIKTPDGLAREMHRTKTSAGTVYAMMADEVYIRASMFESK				
	GSIDWYGVILWELTNGEYVPRGIDGLAVYAGVMNKTLADIPSTCEPPAKIMEDCW				
	NPDHRSRSTNITLIDULTTIDESGFPEFKRDSFICLDDNKKHEIQEAFDOLRAKEKEL				
	RTWELDLRALDQKNOBELRRRQELAEKEIDILEREHLTIHQLCQEPVKKRRK				
	GKFKSRILKLDGNRI SLPSDFORKEQVQASPTMDKRSKLINSRSPASPTIIPRIAR				
	AIQTLPGESSKTWGRSVYVPEEGEEDERKAPKRWGSGTGLQKEIASGDEGSPQ				
	REFRANGKSTPESPHFLGLKSLVDYQKQWSSAPNLYKGRSSPALPGTSLMEMAA				
	LLASWVYDIEEDDESEGGSGESRLOHSPQSYLCIPPRGDDGSPDSGIEHEE				
	PTPNASATSTPOLTPNLSIKRGAAHRRGEPVALLGCCAVYLAATGLGFDILAECQILL				
	PLDEEPRAREEKKRRRGLFORSRSPRSTSPSRKLKKEEPMILAGDBASLITLIS				
	LSSTISECHSTSLRSDDELIVYEMPVSEAPLSPCHNPLVANNRVEREFKDPVN				
	SLTPHYVITTPSQSPSHRRTPSGALKPILYLAIRSPSSNSGLSPSGAQLKTPSPS				

BASE COUNT	875 a	1153 c	1134 g	769 t
ORIGIN				
Query Match	57.4%	Score 94.2;	DB 9;	Length 3931;
Best Local Similarity	73.6%;	Pred. No. 2.7e-17;		
Matches 120;	Conservative 0;	Mismatches 43;	Indels 0;	Gaps 0;
QY	2	CACCGGACATCAGGACGGAATATTTTGGCTACTTGAAGAATAGAACATGATGACATC	61	
Db	641	CACCGGACCTTAATCCAGCAACATATTGATTCCTCCAGAAAGGTGAGAAATGAGACCTG	700	
QY	62	TGCATTAATTAACCTTTAAAGATTAACGATTTTGGCTGTGGGAGGAATGGCACAGGACACC	121	
Db	701	AGCAATTAAGATTCTTAAGATGATCACTGATTTTGGCTGTGGGAGGAATGGCACAGGACACC	760	
QY	122	AAATGAGCAGACGACGACCTTATGCTGATGTGCGCCAGAAAG	164	
Db	761	AAGATGAGTGGCGGACGAGGACGTATGCTTGATGTGACCCGAAG	803	
RESULT 12				
HSU07747	3531 bp	mRNA	linear	PRI 12-JUL-1994
LOCUS	Human SH3 domain-containing proline-rich kinase (sprk) mRNA,			
DEFINITION	complete cds.			
ACCESSION	U07747			
VERSION	U07747.1	GI:464027		
KEYWORDS				
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
AUTHORS	1 (bases 1 to 3531)			
TITLE	Gallo,K.A., Mark,M.R., Scadden,D.T., Wang,Z., Gu,Q. and Godowski,P.J.			
FEATURES	Identification and characterization of SPRK, a novel src-homology 3 domain-containing proline-rich kinase with serine/threonine kinase activity			
JOURNAL	J. Biol. Chem. 269 (21), 15092-15100 (1994)			
MEDLINE	94253068			
PUBMED	8195146			
REFERENCE	2 (bases 1 to 3531)			
AUTHORS	Godowski,P.J.			
TITLE	Direct Submission			
JOURNAL	Submitted (15-MAR-1994) Paul J. Godowski, Genentech, Inc., 460 Point San Bruno Blvd., South San Francisco, CA 94080, USA			
FEATURES	Location/Qualifiers			
SOURCE	1..3531			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/cell_line="CMK11-5"			
	/clone_id="gt 10"			
	1..3531			
	/gene="sprk"			
	451..2994			
	/gene="sprk"			
	/standard_name="SPRK"			
	/standard_name="src-homology 3 (SH3) domain-containing proline-rich kinase"			
	/note="submitter comments: serine/threonine protein kinase, proline-rich, src-homology 3 (SH3) domain, leucine/isoleucine zipper motifs"			
	/codon_start=1			
	/product="serine/threonine protein kinase"			
	/protein_id="AA119647.1"			
	/db_xref="GI:464028"			
	/translation="MEPKSLFLKSPGSMNGSGGGGGGGRPEGSPKAGYANPVTATLFDYEPSCDPLARKDRLVYLSRDAISDEGMAGVGGOVGIPSYVSVSGGPPEVASFDELRLLEVYIGIGFGFYVYSGRGLVYAKARQDDDEISTASGVQERLRAMLAHFNIALKAVCLDEEPMLCLVMEYAAAGPLSRALGAKRPVPHLVNM			
gene				
CDS				

TTONSACSTAMMAEVIATASFSSKSDWSEFIMLLEITGVYQJIDGLALVAYC
 ANKTLITIPSCPPPEALOMDCAMODHRRDPAISLLOOLAEALVQJREMPDSEF
 EYSEMGKEITOEGLDEBAPKREKLTREELTRPARBORSQJOLRBRHLLAOEML
 HEBEELITLLLODVDBERHVRKRTFKSKIRANDGGRISMPDEDFKRTIYOASP
 GLDRRNRYEAVGQSPTEPRRAJOLEAPBQOANGROSPLREDSNERRACVMAW
 GPSPCEAONGRRSRMDEATLIDSDSPSLGSPTEPALANGNPRSLBEEBKEP
 RVPAPERSSSOTPILOALLRGVILASLIGRLOPGGREGESPTTPPPPTPL
 PACCTEPPSPILFSLKTPDPSPPRPAJLLIDGIPVQGRNKSPPREBECGVSL
 PPGTSKRAACGTGPRSPPLGLISNPRPSLKRSDIDPMYSVAGNRPSTLPSQPAR
 RAAPVTLPPDSPPEWDSPPANPGGQJOCRAOTKMGADAPWPELGP"

